

GIS Project: Mapping Limited English Proficiency Individuals

Language Line Services Marketing Department

Submitted by Brian P. Etkin, December 10, 2005

This project applies the principles of geographic information systems to the specific marketing problems facing Language Line Services (LLS), the industry leader in over-the-phone interpretation services (OPI). The primary goal is to develop a tool that enables LLS marketing managers to identify key markets, as well as the dynamics of those markets, for its core products. Analysis draws from publicly available data provided by the United States Census Bureau.

About LLS:

Language Line Services is the industry leader in over-the-phone interpretations services. Its clients include a broad range of industries. In the private sector, LLS serves major telecommunications, airline, finance, insurance and retail companies, allowing its clients to provide quality service to non-English speaking consumers. In the public sector, LLS provides OPI to local, state and federal agencies, ranging from emergency call centers to courts to social service offices. Many agencies are required by federal law to provide some form of interpretation services, and LLS helps them stay in compliance with these regulations. Similarly, many healthcare providers are also bound by law to provide equal access to limited English proficiency (LEP) consumers.

The LLS model is fairly straightforward. An LEP customer contacts a business or agency. This contact may be in person (e.g., at a branch office) or over the phone (e.g., through a customer service call center). The employee, having established the language requirement of the customer, establishes the equivalent of a three-way call with an LLS interpreter. The interpreter then acts as a communications bridge between the service-provider and the LEP customer.

How GIS can Help:

The LLS marketing team has used demographic data for several years now to try to identify target markets. Most of this data has been taken from the United States Census Bureau. The key indicator that they have used is the number of people reporting limited English-speaking ability in the 2000 census. Higher numbers of LEP population would, naturally, seem to indicate more attractive markets.

The question is how does one move from looking at the number to understanding the picture? How can this key indicator be organized in such a way that it reveals critical information? How can LLS utilize public domain demographic information in such way as to increase its competitive advantage in an industry that has become almost cut-throat in its competition?

GIS is an effective means of organizing this type of information. All census data can be represented cartographically with a great level of detail (down to the census tract!). If this, however, were the only application of GIS, then it would be simply a glorified map-making tool that lets us create pretty pictures. GIS is more than maps. It is a tool for analysis and decision-making.

In the case of LLS, GIS could allow marketing managers to layer additional information on top of the language proficiency indicator. This information may be demographic: household income, age, population density. Or, it may be LLS's own information: number of minutes per region.

The advantage of layering this information is that we can begin to examine relationships between variables. We can even isolate certain variables. For example, we could look for communities that have a minimum density of LEP individuals and a median household income over a certain amount.

The LLS marketing department could apply this GIS in several ways. First, GIS could aid LLS in its external marketing efforts. The market insights that LLS gains from this project will allow it to approach potential clients and provide not just interpretation services but a full range of multi-language solutions. For example, LLS could help banks evaluate the language service needs of individual branch offices based on local information.

GIS could potentially aid in internal information analysis. Assuming that the data is available, marketing managers could evaluate market penetration within various geographic regions. LLS could identify regions where the number of LEP individuals is high but the volume of minutes for that region is low. Managers could then decide how best to increase penetration based on other demographic factors.

Geoprocesses and Geoanalysis:

As with every GIS projects, the specific geoanalysis tools applied to LLS's marketing needs depended on the nature of the problem being addressed. Many marketing applications of GIS technology revolve around analysis of demographic and socio-economic data. Our survey of LEP individuals is no different.

This project involved three basic stages. During the first stage, I created the geodatabase upon which the GIS would be based. In the second stage, I visually manipulated the geodatabase to reveal meaningful patterns and trends in the data. The final stage resulted in maps with which decision makers could analyze different options available to them.

Stage 1: The Geodatabase

The heart of the LEP geodatabase is demographic data from the 2000 United States Census. The extended version of the census provides detailed information about the population of the United States. Among the thousands of variables in the extended data set are data about English-language proficiency. This data is disaggregated further by native language groups and age bracket. All these variables are available at differing levels of geographic agglomeration. The broadest geographic region is the entire United States, but data can be found down to the census tract level. The data available in the census are indeed a robust and powerful set of information.

Unfortunately, the very qualities that make the census data set valuable also make it extremely unwieldy. The data are strewn over hundreds of files and dozens of gigabytes in formats especially designed by the Census Bureau. Unlocking the full potential of the census data

requires advanced data mining techniques. Such data manipulation is not for the novice or the faint of heart.

Census data comes in an extremely granular format that is not necessarily useful for analysis. A significant portion of my project has involved manipulating massive sets of raw data from the Census Bureau to derive variables that can be applied to our analysis of Limited English Proficiency populations in the United States. All records reflect data at the ZIP code level for the entire United States. The following are the variables which I have created based on the U.S. Census data:

1. Total Population
2. Total Population, Age 5-17
3. Total Population, Age 18-64
4. Total Population, Age 65+
5. Total Spanish-speaking Population
6. Total Indo-European-speaking Population
7. Total Asian-speaking Population
8. Total Other-speaking Population
9. Non-native Speakers, 5-17
10. Non-native Speakers, 18-64
11. Non-native Speakers, 65+
12. Total Non-native Speakers
13. Total Spanish Speaking, 5-17
14. Total Indo-European, 5-17
15. Total Asian, 5-17
16. Total Other, 5-17
17. Total Spanish, 18-64
18. Total Indo-European, 18-64
19. Total Asian, 18-64
20. Total Other, 18-64
21. Total Spanish, 65+
22. Total Indo-European, 65+
23. Total Asian, 65+
24. Total Other, 65+
25. LEP, 5-17
26. LEP, 18-64
27. LEP, 65+
28. LEP Spanish Total
29. LEP Indo-European Total
30. LEP Asian Total
31. LEP Other Total
32. LEP Spanish, 5-17
33. LEP Indo-European, 5-17
34. LEP Asian, 5-17
35. LEP Other, 5-17
36. LEP Spanish, 18-64
37. LEP Indo-European, 18-64
38. LEP Asian, 18-64
39. LEP Other, 18-64
40. LEP Spanish, 65+
41. LEP Indo-European, 18-64
42. LEP Asian, 65+
43. LEP Other, 65+
44. LEP Total
45. LEP Density (=LEP Total/Total Population)

Stage 2: Geoanalysis

Once the geodatabase was developed, the next step was to associate the demographic data with geographic information. Through several rounds of data manipulation, I was able to associate the census records with five-digit ZIP code regions. Once the records were geocoded, the demographic information could be represented in cartographic form. All maps utilized in this analysis are based on the existing maps of the United States provided from ESRI.

The main objective of geoanalysis for this project was to identify regions of the United States that offered the greatest market opportunity for LLS. GIS allows us to identify those areas by visualizing a given set of demographic characteristics. For LLS, the most obvious characteristic to analyze is the number of LEP individuals in any given ZIP code. The absolute number of LEP residents may not provide an accurate picture of the interpretation needs of a region, which is why I also examined LEP density (the ratio of LEP population to total population). In order to focus attention on the highest priority regions, I also performed a SQL query which selected only those ZIP code areas in which the LEP density exceeded 10 percent of the total population.

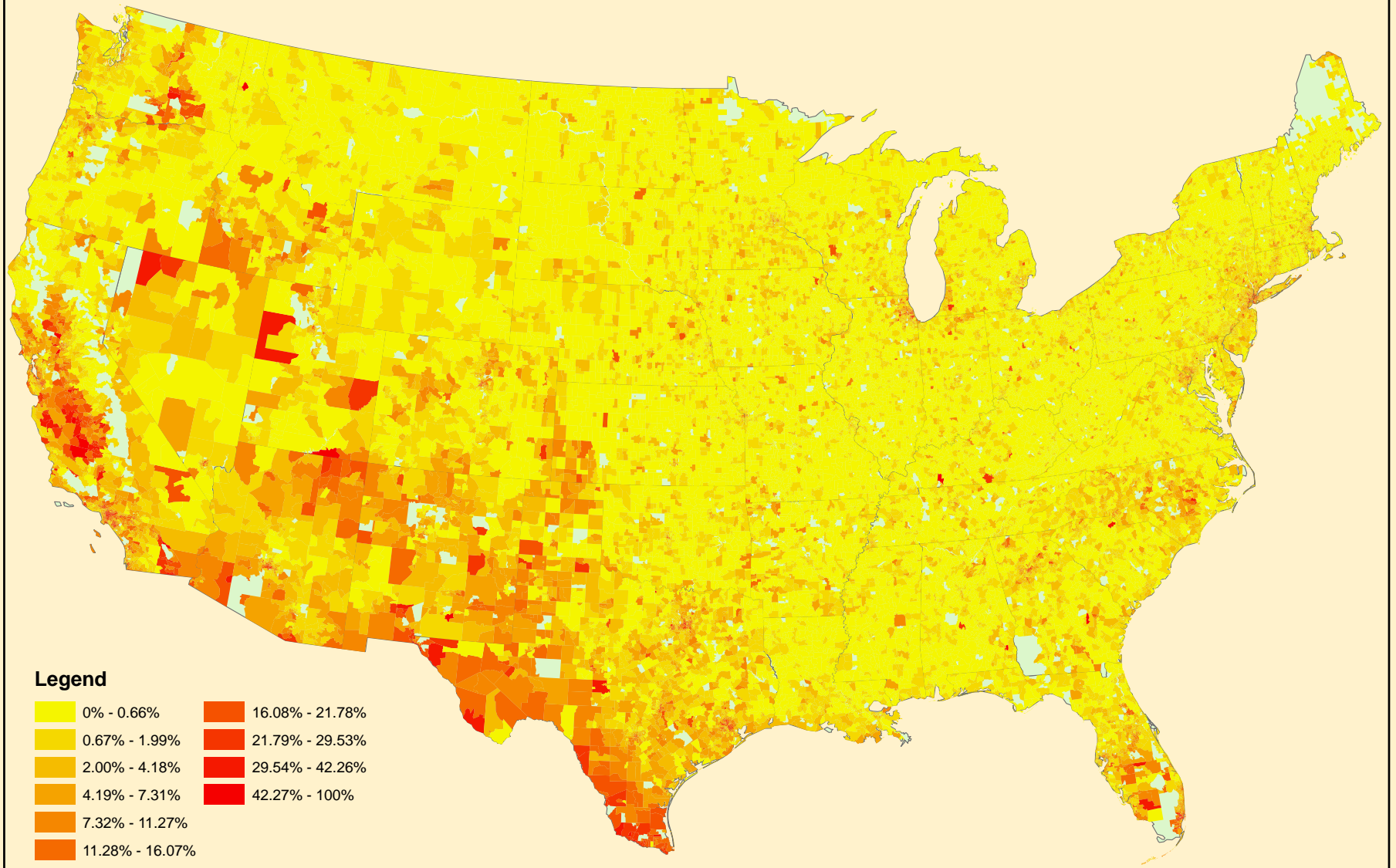
Visualization of LEP population by ZIP code allowed me to identify clusters of high-density LEP regions. When this information is represented cartographically, the eye quickly sees patterns and clusters. Having identified these clusters, I then was able to prepare more detailed analyses of the high-density regions. These analyses included visualization of other variables of LEP population of a cluster, such as age or primary language group.

Stage 3: Presentation

I have produced three final maps. The first shows LEP density for the entire United States. The second shows LEP density for regions where the LEP density exceeds 10 percent. The final map is an example of how more detailed information can be provided for each cluster—in this case six clusters in the State of California.

Concluding Remarks

This project only scratches the surface of how GIS tools can uncover valuable information for marketing managers. An enterprise-level geodatabase could fully realize the potential of GIS and demographic marketing. Such a geodatabase would need constant updating and management, but it would also provide a level of sophistication in analysis that would ensure LLS a sustainable competitive edge over its competitors. A geodatabase of this caliber would not only tap into other census variables, but also take into account in-house metrics. Better information leads to better analysis. Better analysis results in better decisions.



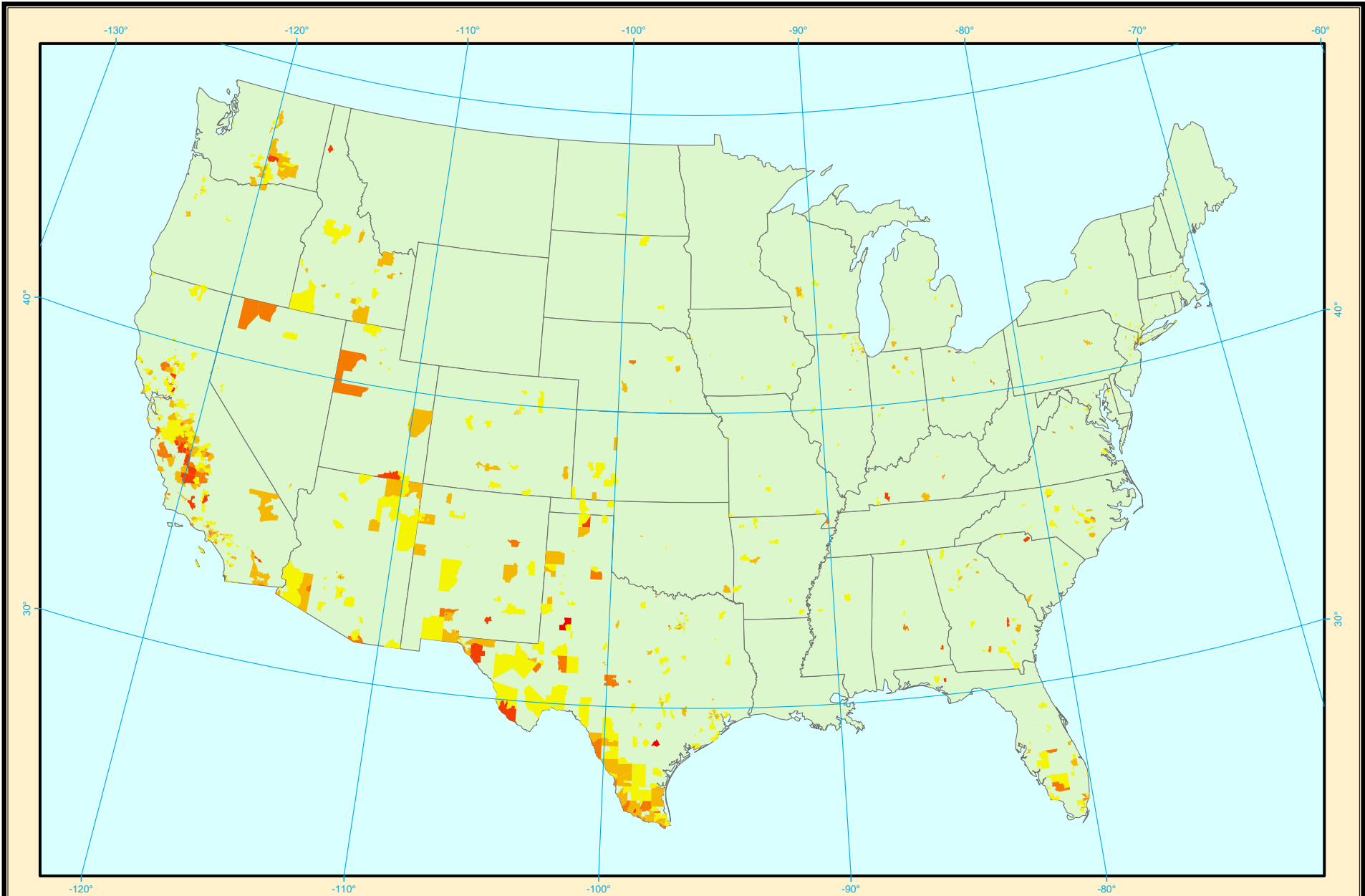
Albers Projection
 Control Meridian: -96
 1st Std Parallel: 20
 2nd Std Parallel: 60
 Latitude of Origin: 40

U.S. ZIPs with LEP Density

Brian P. Ettkin
 November 9, 2005

Source: U.S. Census
 Created in ArcGIS 9 using ArcMap





Legend

LEP Density

- 10% - 16.07%
- 16.08% - 24.35%
- 24.36% - 36.64%
- 36.65% - 59.05%
- 59.06% - 100%


Albers Projection
 Control Meridian: -96
 1st Std Parallel: 20
 2nd Std Parallel: 60
 Latitude of Origin: 40

U.S. ZIPs with LEP Density > 10%

Brian P. Etkin
 November 9, 2005

Source: U.S. Census
 Created in ArcGIS 9 using ArcMap

0 50 100 200 300 Miles



CALIFORNIA

LIMITED

ENGLISH

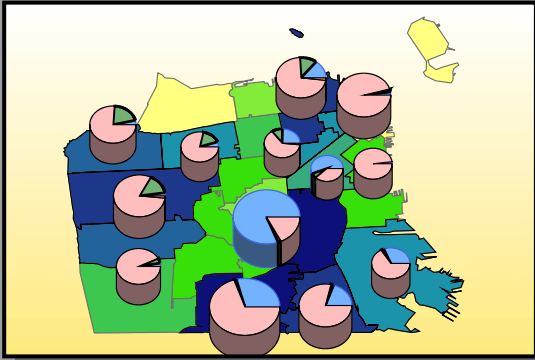
PROFICIENCY POPULATION

These maps represent the Limited English Proficiency (LEP) population residing in the State of California. The state map shows actual resident counts for each California county. Darker shades of blue represent higher absolute numbers.

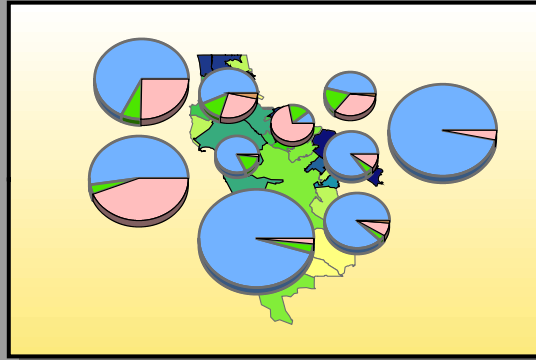
The smaller maps provide greater detail for the regions with higher stocks of LEP residents. Each territory of the smaller maps represents a ZIP code area. Again, darker shades of blue represent higher numbers of LEP residents.

The pie charts show the language group composition of the ZIP code areas with the highest numbers of LEP residents in each region. The possible language groups are Spanish, Indo-European, Asian and Pacific Islander, and Other language groups. The relative size of the pie represents the total

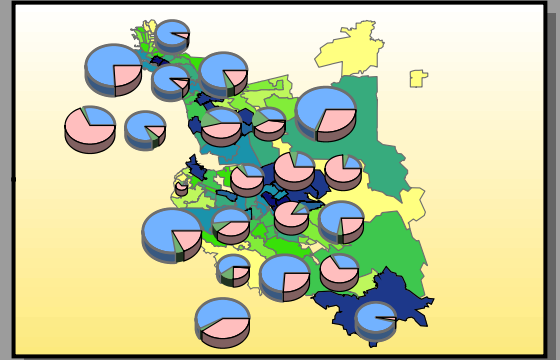
number of LEP persons in each ZIP code area compared with other ZIPs in the region.



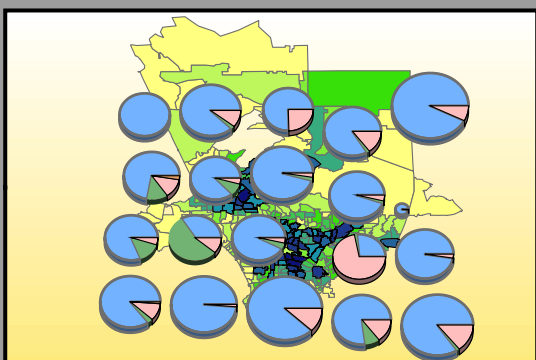
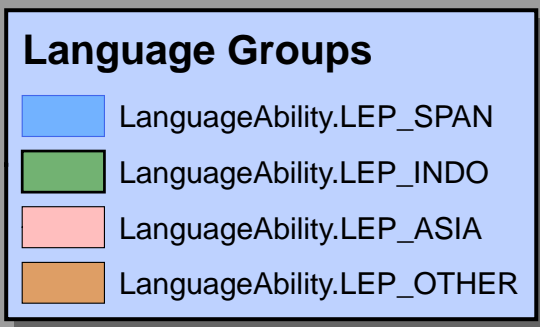
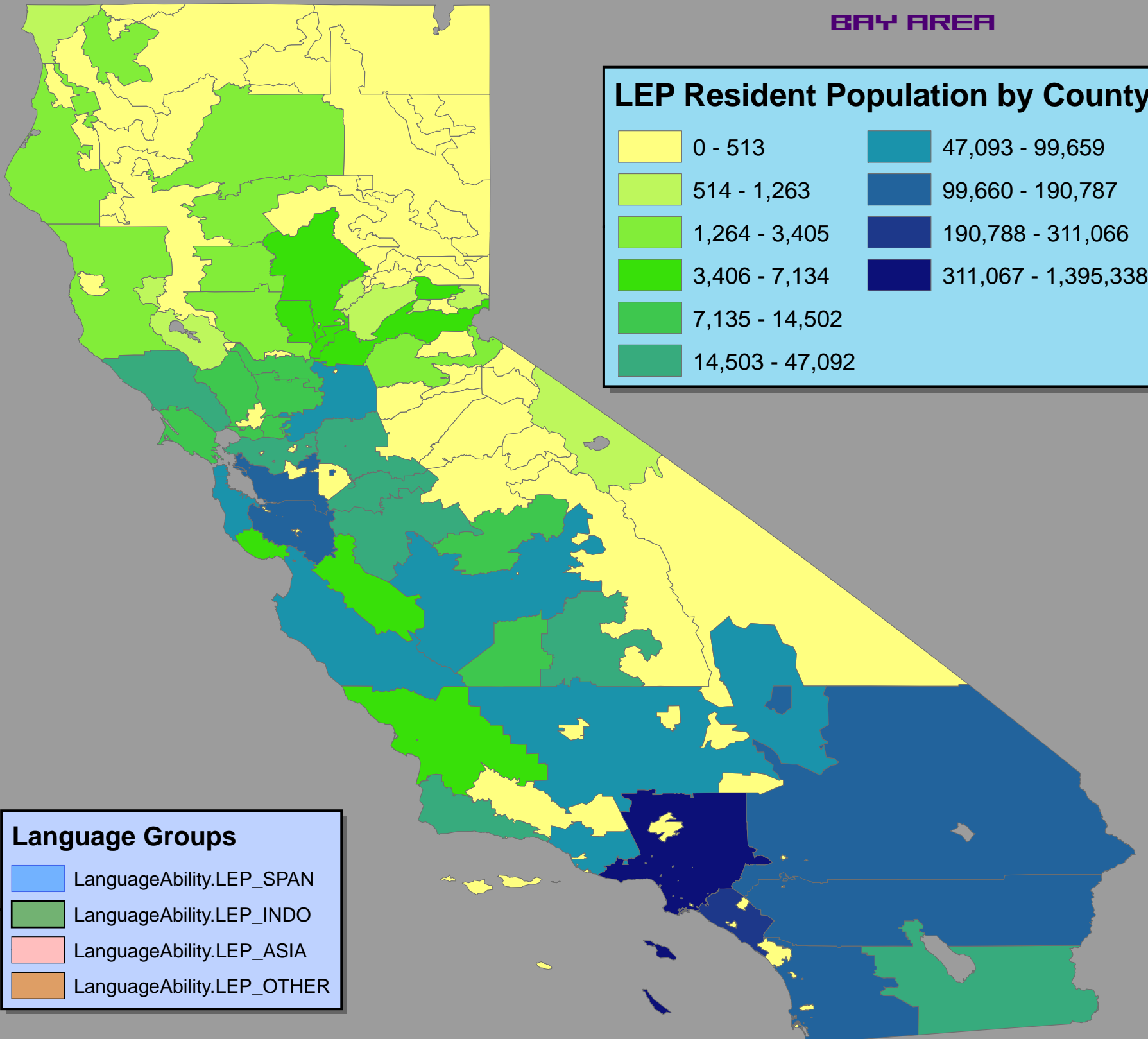
SAN FRANCISCO



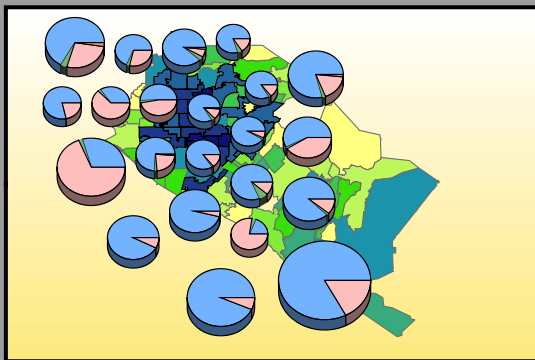
PENINSULA AREA



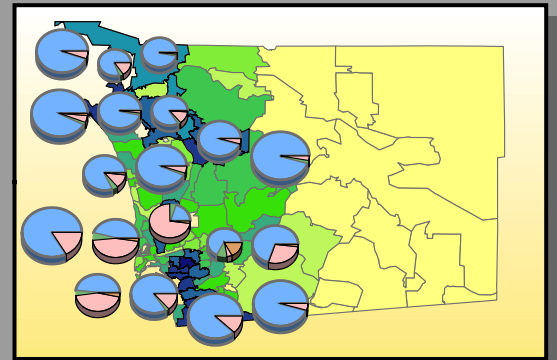
EAST AND SOUTH BAY AREA



LOS ANGELES



ORANGE COUNTY



SAN DIEGO